

Capítulo II

Datos y Metodología

2.1 Datos

2.1.1 Datos de precipitación

Los datos de precipitación observada utilizados para este estudio se obtuvieron de la red pluviométrica del Valle de México, a cargo de la Dirección General de Construcción y Operación Hidráulica (DGCOH) del Distrito Federal. Esta red pluviométrica es la más completa en cuanto a cobertura espacial y temporal sobre la Ciudad de México. Actualmente, la red está compuesta por 78 estaciones pluviométricas (Fig. 9) que registran datos al menos cada 10 minutos.

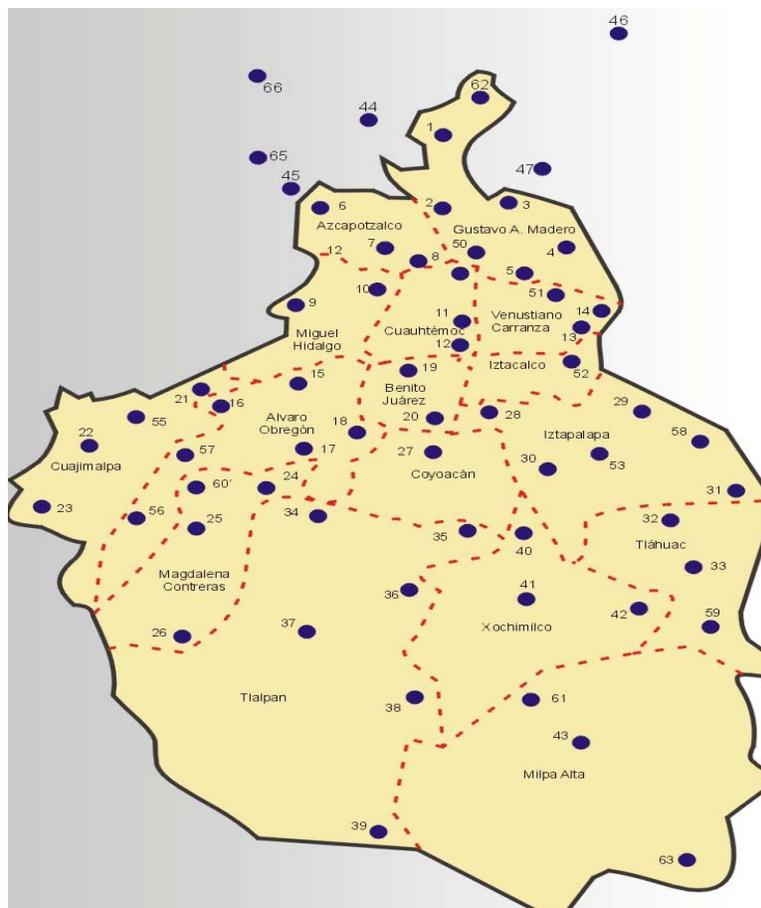


Fig. 9 Distribución de las estaciones pluviométricas de la DGCOH.

Durante episodios de lluvia, los datos son transmitidos en tiempo real. Esto permite, entre otras cosas, analizar con gran detalle la evolución de las tormentas. La continuidad de los registros en esta base de datos permite caracterizar las lluvias de acuerdo a su distribución espacial, intensidad, duración, etc.

2.1.2 Datos de temperatura y vientos

Los datos de vientos, temperatura y humedad son de la Red Automática de Monitoreo Atmosférico (RAMA), dependiente de la Secretaría del Medio Ambiente del Distrito Federal. Esta red cuenta con 32 estaciones automáticas distribuidas en el Distrito Federal y el Estado de México. De éstas, 15 miden parámetros meteorológicos (Fig. 10) que son reportados como media horaria de variables como temperatura, humedad relativa, así como dirección y magnitud del viento.

El periodo que cubre la base de datos, tanto de precipitación como de temperatura, va de mayo de 1993 a octubre de 2002. Los datos de las estaciones, distribuidas irregularmente, son utilizados para generar campos diarios en mallas, pues es más práctico trabajar con arreglos matriciales al momento de comparar pronósticos y observaciones.

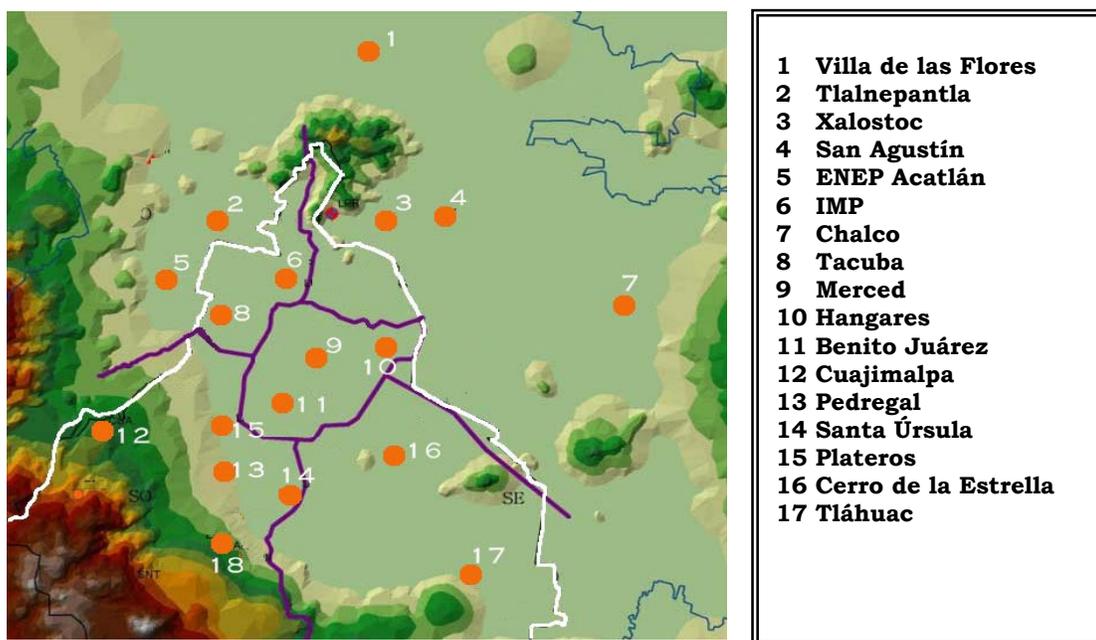


Fig. 10 Distribución de las estaciones meteorológicas de la RAMA en el Distrito Federal y sus alrededores.

Para generar los campos observados, se utilizó un esquema de interpolación del tipo Akima (Akima, 1978). Así, se generaron mallas de datos con resolución espacial de 10 km, que es equivalente a la distancia media entre estaciones.

2.2 El modelo de mesoescala MM5

Uno de los modelos de área limitada más populares es el MM5. La quinta generación del MM5 (MM5-PSU/UCAR, por sus siglas en inglés) es el resultado de diversos desarrollos en la Universidad de Pennsylvania (Warner y Anthes, 1978), y en el *National Center for Atmospheric Research* (NCAR, por sus siglas en inglés). Por ser un modelo de dominio público para la comunidad de las Ciencias Atmosféricas, sus cambios y adecuaciones son probados por numerosos usuarios. Es por ello que el MM5 se ha convertido en un modelo cada vez más confiable y robusto.

Dentro de las propiedades del MM5 versión 2 se encuentran:

- Diversos tipos de proyección (Polar estereográfica, Lambert Conformal, y Mercator).
- Esquemas de dinámica hidrostática o no-hidrostática, lo que permite realizar experimentos en escalas de unos cuantos kilómetros.
- Consideraciones de las características del terreno (uso de suelo, topografía, cuerpos de agua, etc.)
- Capacidad de asimilación de datos observados (radares, radiosondeos, superficie, etc.)
- Esquemas avanzados de la microfísica de nubes; parametrizaciones de cúmulus; esquemas de capa límite planetaria y procesos de radiación.
- Flexibilidad para realizar simulaciones para más de un dominio, ya sean anidados al dominio madre o subdominios en diferentes regiones.
- Puede ser ejecutado en varias plataformas de cómputo, que van desde supercomputadoras hasta computadoras personales con sistema *LINUX*.
- Facilidad de acoplamiento con otros modelos de mesoescala.

El MM5 es un sistema de modelación compuesto por módulos de preprocesamiento y postprocesamiento de datos con una estructura como la mostrada en la Fig. 11, y que se describen a continuación.

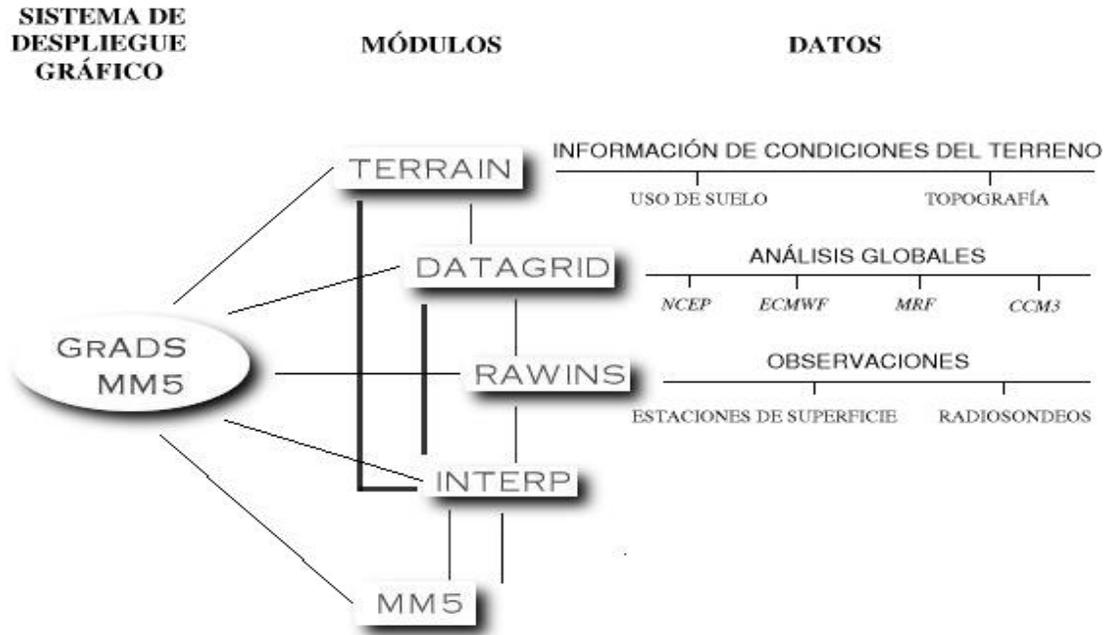


Fig. 11 Esquema de la estructura del modelo de mesoescala MM5.

TERRAIN: Preprocesamiento que define los dominios madre y anidado, la resolución y el tipo de proyección del mapa. Además, asigna los valores de la topografía y las características del terreno.

DATAGRID: Preprocesamiento con el que a partir de los datos de gran escala de viento, humedad, temperatura, etc., (campos iniciales) se construyen las mallas de los dominios a trabajar con la resolución deseada, generando así análisis de datos interpolados con alta resolución espacial, listos para ser procesados por los módulos de RAWINS o INTERP. La información utilizada puede ser histórica, para estudios de diagnóstico o simulación, o en tiempo real para pronóstico operativo del tiempo.

LITTLE_R: Preprocesamiento donde se realiza un análisis objetivo de la información utilizando datos observados en superficie de estaciones meteorológicas y de altura, como radiosondeos, globo piloto, etc. La asimilación de datos tiene la finalidad de mejorar los campos de las variables utilizados como condición inicial.

INTERP: Preprocesamiento donde se interpola de los datos iniciales de coordenadas de presión a coordenadas verticales sigma. Este tipo de conversión de coordenadas permite una mejor representación de la orografía (Yong-Fu y Zhong, 1987).

MM5: Este módulo resuelve numéricamente un sistema de ecuaciones primitivas en términos de flujos en coordenadas sigma por medio de diferencias finitas, usando el método semi-implícito (Haltiner y Williams, 1979) sobre una malla numérica Arakawa tipo-B¹ (Fig. 12).

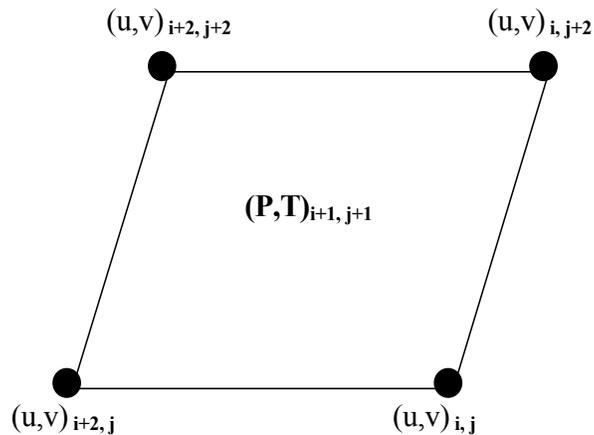


Fig. 12 Malla Arakawa tipo-B.

¹ Este esquema define a las variables escalares (p,t, etc) en el centro de la malla cuadrada (centrada sobre el dominio), mientras que las componentes u y v del viento están colocadas en las esquinas, coincidentes con los puntos de malla del dominio.

2.3 Datos meteorológicos para el MM5

Un modelo de pronóstico numérico del tiempo requiere de condiciones iniciales y de frontera para realizar simulaciones o predicciones a 12, 24, 36 o 48 horas. En pronósticos operativos, las condiciones inicial y de frontera (9 condiciones en intervalos de 6 hr) fueron obtenidas de los pronósticos a escala global realizados diariamente en el Departamento de Aviación de los Estados Unidos² (AVN, por sus siglas en inglés), desarrollados en conjunto con el *National Center for Environmental Prediction* (NCEP, por sus siglas en inglés).

Los datos de AVN se generan en mallas globales con una resolución de 1° x 1°. Las variables necesarias para el MM5 son:

- Componente zonal y meridional (u y v) del viento
- Temperatura
- Humedad relativa
- Presión reducida al nivel del mar
- Altura geopotencial

Todas las variables se obtuvieron en 10 niveles: 1000, 850, 700, 600, 500, 400, 300, 200, 150 y 100 mb, con excepción de la humedad relativa que sólo se reporta hasta el nivel de 300 mb. De la misma forma, se obtuvieron los datos medios semanales de la temperatura superficial del mar (con resolución espacial de 1° x 1°), contruidos por medio de métodos óptimos de interpolación por el NCEP² (Reynolds y Smith, 1994), los cuales sirven para actuar como condición de frontera en áreas oceánicas.

Para zonas continentales, la topografía y el uso de suelo son fundamentales para el MM5 como condición de frontera, pues en gran medida son los parámetros responsables de generar circulaciones de mesoescala.

¹ http://wesley.web.noaa.gov/ncep_data/index_sgi51_png.html.

² <ftp://ftp.ncep.noaa.gov/pub/cmb/sst/oiupdate>

2.4 Metodología

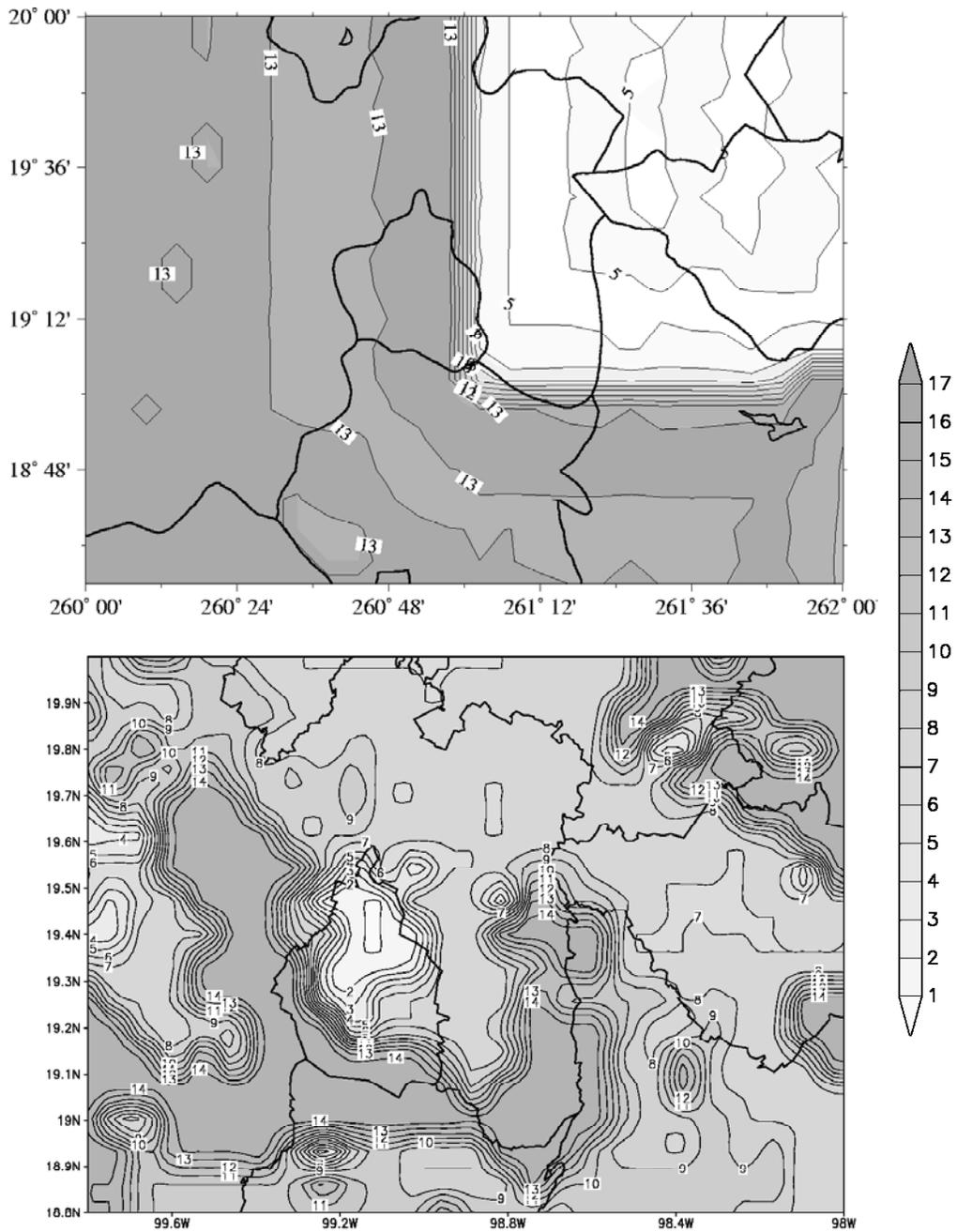
Durante 3 años (2000, 2001 y 2002) en el CCA-UNAM se han realizado pronósticos operativos con el MM5 para el Valle de México, obteniéndose un banco de datos de las salidas del modelo. Para evaluar los resultados de tales pronósticos y estimar cuál es la resolución espacial más adecuada a utilizar con fines de predicción del tiempo, se utilizaron 3 distintas resoluciones en el dominio madre: 24, 45 y 60 km, y en el anidado: 8, 15 y 20 km, respectivamente. 8 km. es la máxima resolución que dinámicamente puede utilizarse debido al uso de las parametrizaciones de nubes cúmulus. De incrementarse la resolución espacial, las parametrizaciones de ciertos procesos físicos pierden validez, pues sería necesario simular explícitamente la dinámica de las nubes. La evaluación sobre la resolución más adecuada se basó principalmente en la similitud en las simulaciones de los campos de precipitación, por ser esta la variable de mayor relevancia en la región. Sin embargo, sólo en contadas ocasiones se ha analizado la calidad de la simulación de temperaturas y vientos. (Salas, 2000; García, 2002).

Otra opción sería comparar los reportes de las estaciones con un valor obtenido por interpolación de la malla de salida del MM5. Sin embargo, el mayor interés se tiene en la capacidad para reproducir patrones espaciales del tiempo meteorológico.

2.4.1 Criterios utilizados en el experimento de control del modelo MM5.

En el presente trabajo, se consideraron tanto la versión 2 como la 3 del MM5. La diferencia entre ambas versiones ha sido descrita por Dudhia et al. (1999), y radica esencialmente en la mejoría de los datos de uso de suelo y de la topografía (Fig. 13), así como la adecuación para la asimilación de datos, permitiendo simular con mayor confianza la condición inicial.

Dada la amplia variedad de esquemas de parametrizaciones físicas con las que cuenta el modelo, es importante fijar esquemas, denominados como *de control*, para realizar las simulaciones del tiempo.



CATEGORÍAS DE USO DE SUELO:

1. Área urbana, 2. Área de agricultura, 3. Prados, 4. Bosque 5. Bosque de coníferas, 6. Bosque y cuerpos de agua, 7. cuerpos de agua, 8. Pantano, 9. Desierto, 10. Tundra, 11. Hielo 12. Bosques tropicales o subtropicales, 13. Savana. De la 14 a la 17 no hay clasificación precisa.

Fig. 13a Uso de suelo utilizado en el MM5 en la versión 2 (arriba) y versión 3 (abajo).

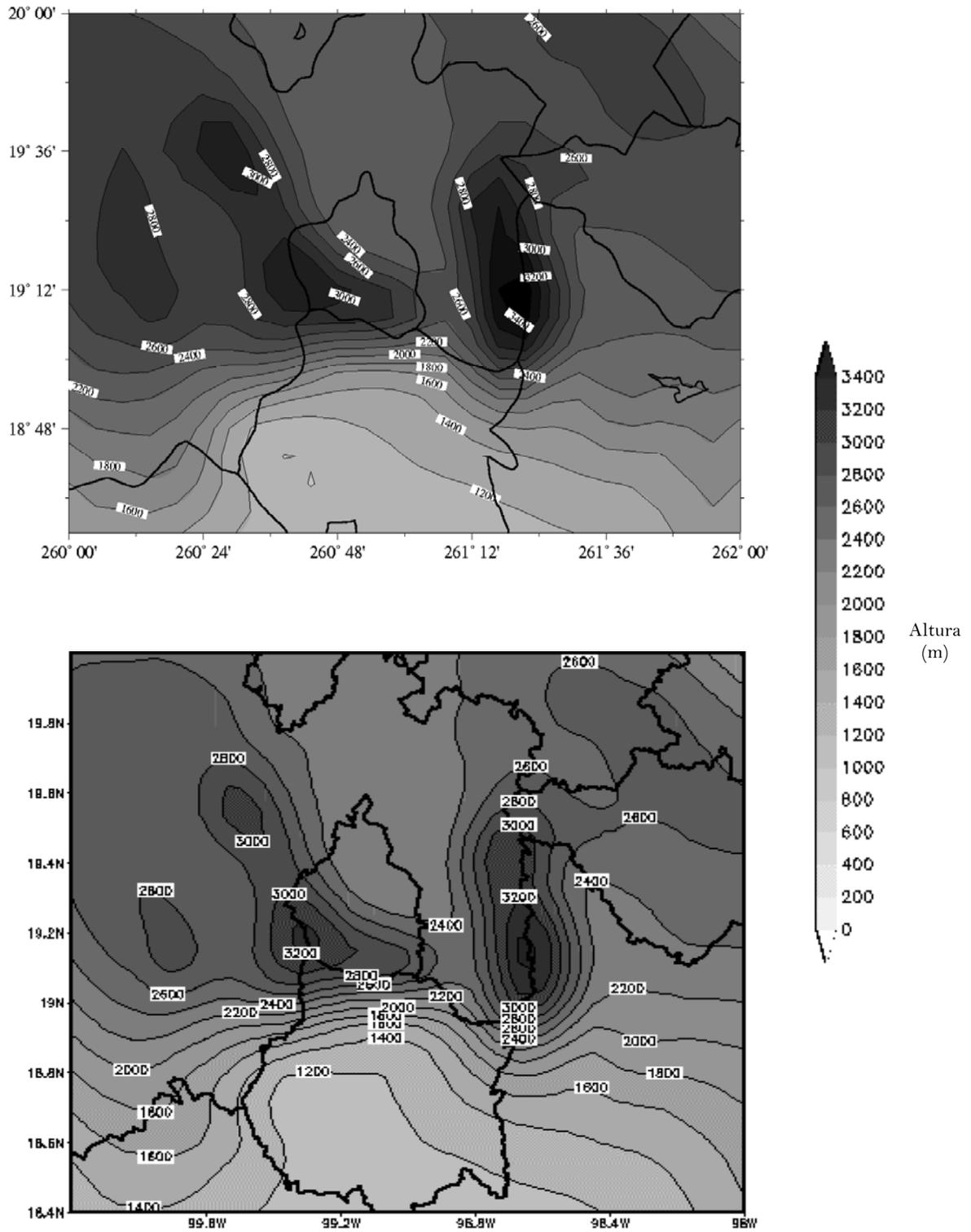


Fig. 13b Topografía utilizada en el MM5 versión 2 (arriba) y versión 3 (abajo).

Las parametrizaciones consideradas en ambas versiones fueron las siguientes:

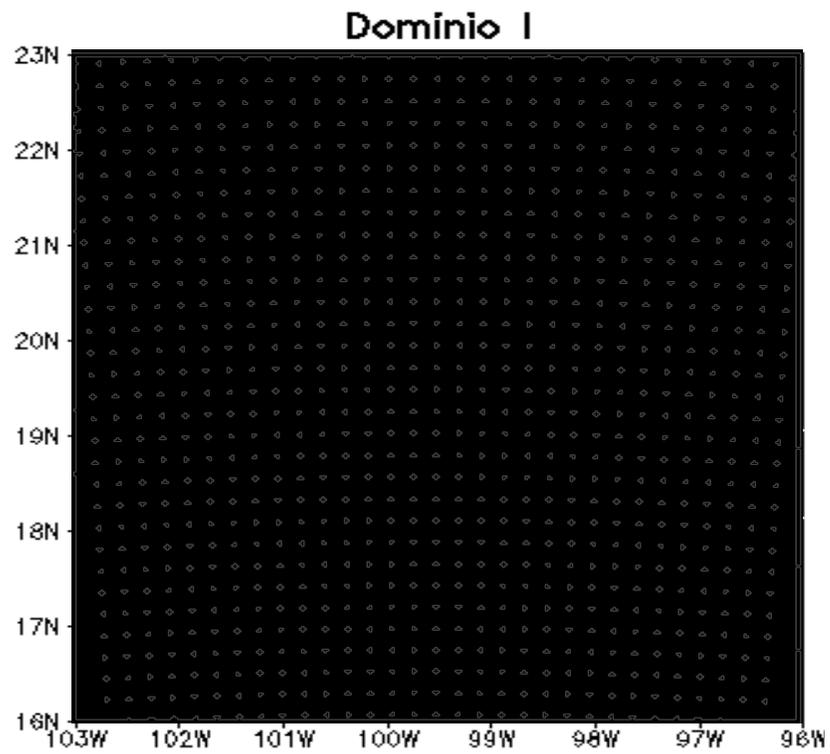
- 2 mallas anidadas interactuando entre sí, siendo tanto la malla madre como la anidada de 31 X 31 puntos cada una, en proyecciones tipo Lambert (Fig. 14).
- La malla madre está centrada en 19.1 N y 99.2 W. La resolución espacial para los dominios madre es de 24, 45 y 60 km; y de 8, 15 y 20 km para los anidados correspondientes.
- Corrida en condición no-hidroestática, utilizando 10 niveles de presión (1000, 850, 700, 600, 500, 400, 300, 200, 150 y 100 mb) interpolados a 23 niveles en coordenadas verticales sigma. La coordenada sigma está definida como:

$$\sigma = \frac{p - p_t}{p_s - p_t},$$

donde p_s es la presión de superficie, p_t es la presión en la cima (100 mb). La coordenada sigma es adimensional y decrece monótonamente con la altura. Mediante el uso de las coordenadas sigma, se puede tener una representación más suavizada de la topografía de la región de estudio.

- Esquema de parametrización de nubes cumulus tipo Kain-Fritsch en ambas mallas. Este esquema permite el mezclado entre nubes para determinar el *entrainment-detrainment*, removiendo toda la energía de flotabilidad disponible (Kain y Fritsch, 1993).
- Esquema de capa límite planetaria desarrollada por Blackadar (1962), caracterizado por una capa de mezcla lineal en la capa superficial.
- Esquema de radiación atmosférica de enfriamiento simple, donde se supone que la razón de enfriamiento atmosférico sólo depende de la temperatura.

a)



b)

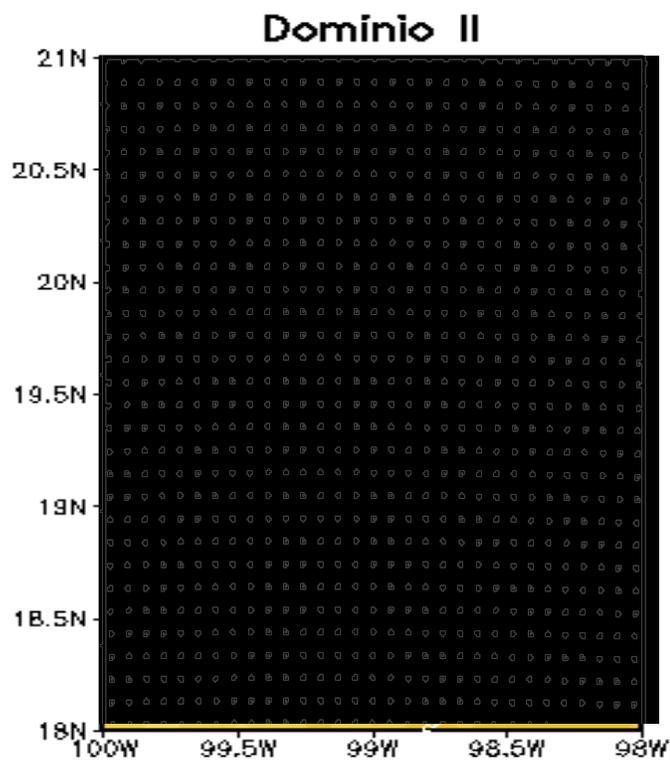


Fig.14 Dominio a) madre y b) anidado en puntos de malla con resolución de 24 y 8 km utilizados en el MM5.

2.4.2 Consideraciones para el caso de la Cd. de México

Para la obtención de mejores simulaciones a corto plazo, es muy importante contar con una buena condición inicial. En la versión 3 del MM5, la asimilación de los datos de superficie y altura se mejoran las condiciones iniciales de campos meteorológicos usando reportes de temperatura, humedad relativa y magnitud y dirección del viento, mediante un esquema de análisis objetivo tipo Cressman (Cressman, 1959) (Fig. 15). Cabe señalar que el proceso de la asimilación de datos no sólo mejora la condición inicial, sino el desempeño del modelo en general. Esto se sigue del hecho de que el pronóstico numérico del tiempo a corto plazo es esencialmente un problema de condición inicial.

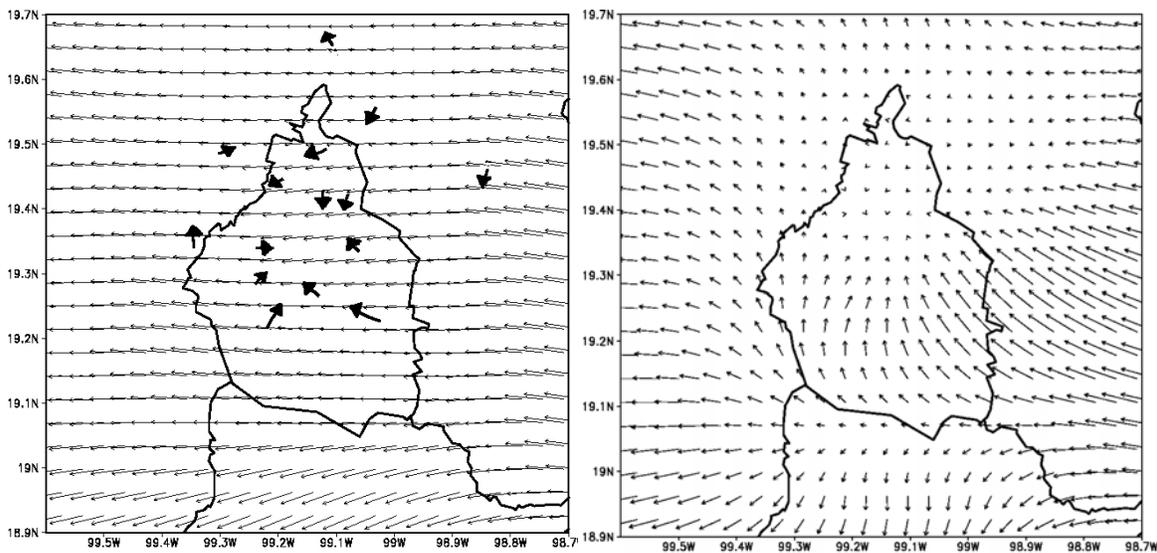


Fig. 15. Comparación del patrón del viento para el día 29 de julio de 2002: a) sin asimilar y b) asimilado en la condición inicial (6 hora local). Los vectores gruesos corresponden a observaciones de la RAMA.

El proceso de asimilación de datos consiste en una serie de pasos entre los que destacan el control de calidad, filtrado de datos equivocados, suavizamiento e interpolación de los datos observados. En este estudio, se consideró un nivel de tolerancia (error) promedio de 30 % en el control de calidad entre las observaciones y los campos de las condiciones iniciales del modelo AVN. En el esquema del análisis objetivo, se utilizó el de tipo Cressman (Fig. 16), el cual crea patrones a partir de una condición inicial modificada por datos distribuidos aleatoriamente. Este esquema está basado en un radio de influencia anisotrópico para modificar los valores de puntos de malla alrededor de cada observación.

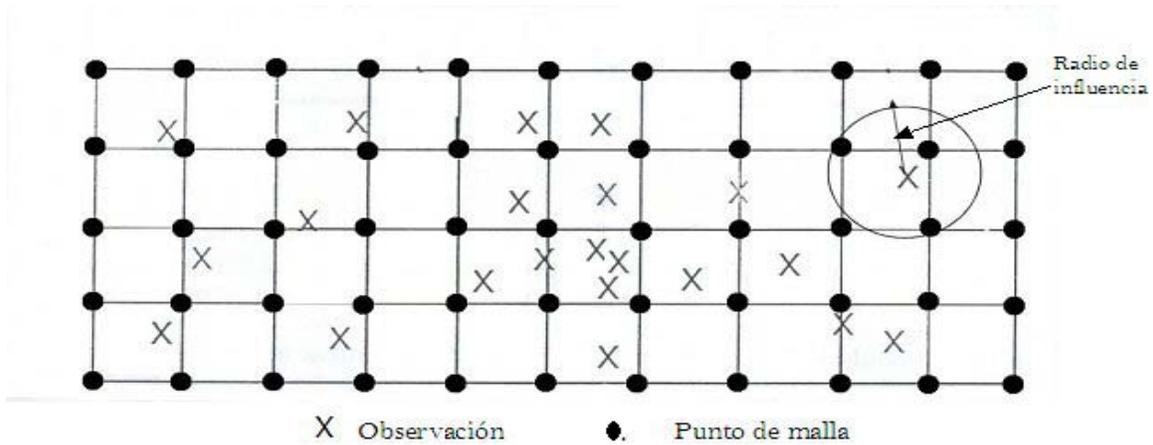


Fig. 16 Esquema de análisis objetivo tipo Cressman. Los puntos representan el arreglo en malla, mientras que las cruces son los datos observados distribuidos aleatoriamente.

Dada la densa red de estaciones de mediciones atmosféricas en el área del Distrito Federal, sólo los datos de la salida del modelo del dominio anidado son comparados con valores interpolados a puntos de malla para su evaluación. Lo anterior implica que los valores observados se construyen con la misma resolución de la salida de modelo su adecuada comparación.

2.5 Esquemas y estrategias de evaluación de pronósticos utilizados

La evaluación del MM5 se puede realizar usando los reportes de estación o análisis en puntos de malla generados a partir de las observaciones, como el campo a reproducir por el modelo.

En el presente estudio, las evaluaciones consideraron 78 estaciones o puntos sobre el Distrito Federal que corresponden a la red de la DGCOH (como en la Fig. 9). En la evaluación puntual, los datos del MM5 son interpolados a los puntos donde se localizan las estaciones mediante un esquema de interpolación bilineal pesado. Este esquema considera los cuatro puntos esquina más cercanos y la distancia de cada uno de ellos al punto de interés, de tal forma que es posible obtener el dato simulado para un punto de interés y realizar comparaciones con lo observado.

Para fines de comparación, los reportes de precipitación y temperatura de las estaciones se interpolaron a una malla con una resolución espacial de 10 km mediante con un esquema de interpolación Akima (Akima, 1978) que es básicamente de *spline cúbicos*. Posteriormente, a partir de las comparaciones puntuales, se crearon campos tanto observados como pronosticados concentrados sólo para el área urbana del Distrito Federal, pues hacia las montañas no hay estaciones y por lo tanto, el campo “observado” es poco confiable.

2.5.1 Evaluación de la temperatura

Para evaluar las predicciones de temperatura de superficie a 24 horas con el MM5, se ha tomado el valor del nivel sigma más cercano a la superficie que está aproximadamente 30 metros sobre el nivel del suelo, interpolado a un punto de interés particular y comparado con el reporte de la estación. Los análisis se concentran en predicciones de temperatura máxima y mínima, así como en las variaciones diurnas de la temperatura.

Como una medida de la calidad de las predicciones con modelos numéricos, comúnmente se utiliza el parámetro estadístico llamado *Raíz del error cuadrático medio* (*rmse*, por sus siglas en inglés) (Wilks, 1995). El *rmse* que se define como:

$$rmse = \sqrt{\frac{1}{M} \sum_{m=1}^M (Y_m - O_m)^2} \quad , \text{ donde:}$$

M es el número total de datos

Y_m es el dato observado

O_m es el dato pronosticado

El uso del *rmse* permite estimar el orden de magnitud de las diferencias entre el modelo y las observaciones, sobre todo en el dominio de interés. Tal esquema ha sido utilizado para pronósticos climáticos de temperaturas (Wilks, 1995).

2.5.2 Evaluación de la precipitación

Al igual que la evaluación de la temperatura, para la precipitación se hace uso del *rmse*. Sin embargo, para el presente estudio se ha diseñado un esquema de evaluación de la lluvia pronosticada, considerando la información de mayor relevancia y siguiendo tres criterios con qué estimar la cantidad de precipitación:

- i) llueve o no llueve,
- ii) cantidad de lluvia acumulada en 24 hrs. (+/- 5 mm), y
- iii) ocurrencia de un evento extremo de precipitación.

Las características de esta evaluación para algunos puntos (Fig. 17) se describen a continuación.

Sea *pred* el pronóstico de lluvia acumulada a 24 hr generado por el MM5, en malla e interpolado al punto de interés, y sea *obs* el reporte puntual registrado en dicha estación.

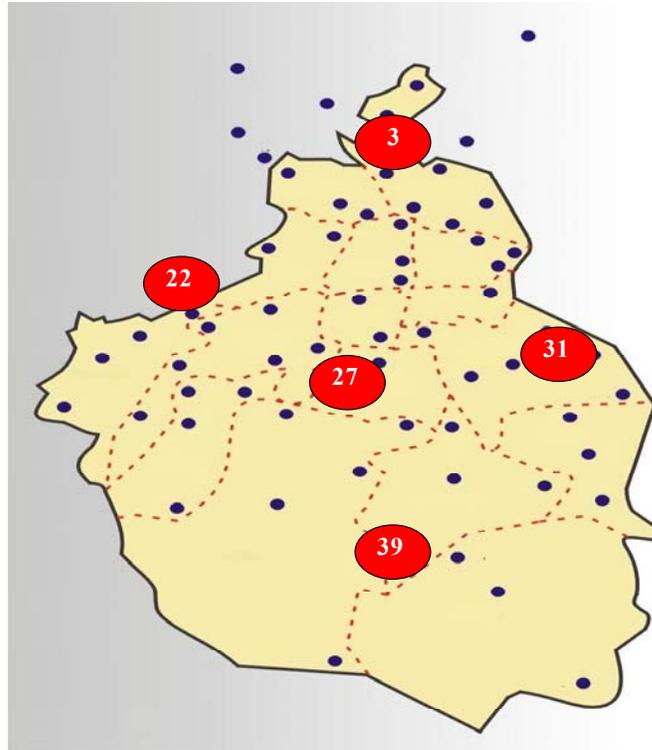


Fig. 17. Estaciones meteorológicas de la DGCOH. Los círculos corresponden a la estación 3-Lindavista, 22-La Venta, 27-Xotepingo, 31-Santa Catarina y 39-Caseta Forestal, utilizadas para las comparaciones puntuales.

Score es el puntaje que se le da al pronóstico. El máximo de puntos que puede alcanzar un día de pronóstico en un lugar dado es de 3, y se da si se pronostica correctamente bajo el siguiente criterio:

- i) criterio: llovió o no llovió:
i.e. si $obs = 0$ y $pred < 1$ mm, es decir, no llueve, entonces $score = 1$, o también
si $obs \neq 0$ y $pred > 1$ mm, es decir, llueve, entonces $score = 1$
- ii) criterio: Cuanto llovió con un margen de error de ± 5 mm alrededor del reporte, i.e.
si $obs \neq 0$, y $obs - 5 \text{ mm} < pred < obs + 5 \text{ mm}$, entonces $score = 1$

- iii) criterio: pronóstico de evento extremo de lluvia
i.e. si $obs \neq 0$, con $obs \geq$ valor de lluvia extremo, y
si $pred >$ valor de lluvia extremo, entonces $score = 1$

Los resultados se interpretan como:

- si el $score=1$, se acertó a la ocurrencia o no de lluvia.
- Si $score=2$, se acertó en el pronóstico de ocurrencia de lluvia y en el rango de lo precipitado con un error de +/- 5 milímetros en 24 horas.
- Si $score=3$, indica acierto en lluvia, en el rango llovido, y acierto al pronosticar un evento extremo.

Mediante este procedimiento la calificación máxima es 3 si llovió y además fue un evento extremo. Los puntos obtenidos por el modelo durante un periodo en particular (e.g. un mes) da información sobre la calidad del modelo y sobre la dificultad de pronosticar cuando se analizan las características de la lluvia. Por ejemplo: un $score$ de 30 puntos en 30 días puede significar que nunca llovió, y el modelo acertó en cada caso. Si el porcentaje de acierto es alto cuando los puntos por obtener fueran muchos (e.g. 90 puntos en un mes), es indicativo de calidad y confianza en el modelo. Sin embargo, cuando el porcentaje es alto y el número de puntos es bajo (e.g. 30 puntos en un mes), se tienen indicios de que el modelo lo hace bien cuando el reto es mínimo.

Asimismo, regiones donde el puntaje por alcanzar sea alto indican zonas en donde ocurren procesos que hacen del pronóstico algo complicado. En otras palabras, serán regiones donde los pronósticos de lluvia diaria serán difíciles de estimar.

Por lo general, el puntaje por alcanzar tiene relevancia en temporada de lluvias, y rara vez alcanzará los 90 puntos como meta (considerando 3 puntos diarios como máximo). Este esquema permite analizar hasta qué punto somos capaces de dar información a un tomador de decisiones, es decir, sobre los alcances y limitaciones de pronósticos a 24 horas.

En la práctica, un meteorólogo operativo deberá utilizar otros elementos descritos en estudios de pronósticos cuantitativos de precipitación (QPF, por sus siglas en inglés), que consideran el radiosondeo y los índices que de él se generan (Corner, et. al., 1999). Dicho trabajo toma en cuenta el pronóstico numérico y otros elementos, como las características del radiosondeo.